Package: mded (via r-universe)

September 9, 2024

Type Package

Title Measuring the Difference Between Two Empirical Distributions

Version 0.1-2

Date 2015-04-27

Author Hideo Aizaki

Maintainer Hideo Aizaki <azk-r@spa.nifty.com>

Description Provides a function for measuring the difference between two independent or non-independent empirical distributions and returning a significance level of the difference.

License CC0

NeedsCompilation no

Repository CRAN

Date/Publication 2015-04-27 11:04:01

Contents

`mded-package` *Measuring the difference between two empirical distributions*

Description

The package provides a function for measuring the difference between two independent or non-independent empirical distributions and returning a significance level of the difference.

Acknowledgments

I would like to thank Professor Gregory L. Poe for his kindness.

Note

Recommended citations:

Aizaki H (2014). **mded**: Measuring the difference between two empirical distributions, R package version 0.1-1. URL http://CRAN.R-project.org/package=mded.

Poe GL, Giraud KL, Loomis JB (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, **87**, 353–365.

Poe GL, Welsh MP, Champ PA (1997). Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. *Land Economics*, **73**, 255–267.

Author(s)

Hideo Aizaki

References

Poe GL, Giraud KL, Loomis JB (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, **87**, 353–365.

Poe GL, Severance-Lossin EK, Welsh WP (1994). Measuring the difference (X - Y) of simulated distributions: A convolutions approach. *American Journal of Agricultural Economics*, **76**, 904–915.

Poe GL, Welsh MP, Champ PA (1997). Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. *Land Economics*, **73**, 255–267.

`mded` *Measuring the difference between two empirical distributions*

Description

The function measures the difference between two independent or non-independent empirical distributions and returns a significance level of the difference.

Usage

```
mded(distr1, distr2, detail = FALSE, independent = TRUE)

## S3 method for class 'mded'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

distr1	A vector of empirical distribution. distr1 is greater than distr2.
distr2	A vector of empirical distribution.
detail	If TRUE, a vector of the difference between distr1 and distr2 is returned.
independent	Set as FALSE when distr1 and distr2 are not independent of each other.
x	An object of S3 class 'mded.'
digits	A number of significant digits.
...	Arguments passed to the function print.

Details

The function measures the difference between two independent or non-independent empirical distributions and returns a significance level of the difference on the basis of the methods proposed by Poe et al. (1997, 2005). Such calculations are frequently needed in empirical econometric studies wherein (marginal) willingness-to-pay distributions that are estimated using contingent valuation methods or discrete choice experiments have to be compared to each other.

Let us assume that X and Y are empirical distributions, which are depicted by the vector $\mathbf{x}$ = (x1, x2, ..., xm), and $\mathbf{y}$ = (y1, y2, ..., yn). The null hypothesis (H0) is X - Y = 0, while the alternative hypothesis (H1) is X - Y > 0. When X and Y are independent of each other, the complete combinatorial method (Poe et al. 2005) provides the one-sided significance level of H0 that is calculated by #{xi - yj <= 0} / m * n, where #{*cond*} provides the number of times that *cond* is true. When X and Y are not independent of each other, the paird difference method (Poe et al. 1997) provides the one-sided significance level of H0 that is calculated by #{xi - yi <= 0} / m, where m is equal to n.

Note that the function may take quite long, and would require large amount of memory to calculate the difference between two *independent* distributions if the argument detail is set as TRUE because the resulting difference is stored as a vector. For example, when distr1 and distr2 each contain 10,000 elements (observations), the vector of the difference contains 100,000,000 elements. If memory is lacking, R would stop running the function, showing an error message related to memory limitaion.

Value

stat	One-side significance level of the difference between distr1 and distr2.
means	A vector of mean values of distr1 and distr2.
cases	A vector of integer values describing a number of cases wherein the *cond* is true and that is false.
distr1	A vector assigned to distr1.
distr2	A vector assigned to distr2.
distr.names	A vector of the names of objects assigned to distr1 and distr2.
diff	A vector of the difference. If detail = TRUE, it is returned.

Author(s)

Hideo Aizaki

References

Poe GL, Giraud KL, Loomis JB (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, **87**, 353–365.

Poe GL, Severance-Lossin EK, Welsh WP (1994). Measuring the difference (X - Y) of simulated distributions: A convolutions approach. *American Journal of Agricultural Economics*, **76**, 904–915.

Poe GL, Welsh MP, Champ PA (1997). Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. *Land Economics*, **73**, 255–267.

Examples

```
set.seed(123)
x <- rnorm(100, 3)
y <- rnorm(100, 1)

out <- mded(distr1 = x, distr2 = y, detail = TRUE)
out
```

Index